DOCUMENT RESUME

ED 220 513 TM 820 529

AUTHOR Humes, Anne
TITLE Writing Samples, Scoring Keys, and Scoring Guides: A Pilot Study.
INSTITUTION Southwest Regional Laboratory for Educational Research and Development, Los Alamitos, Calif.
REPORT NO SWRL-TN-3-79-05
PUB DATE 31 May 79
NOTE 43p.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Elementary Education; *Evaluation Criteria; Guidelines; Pilot Projects; *Scoring; *Test Construction; Testing Programs; Test Items; *Writing Evaluation; *Writing Skills
IDENTIFIERS *Los Angeles Unified School District CA

ABSTRACT

A pilot study was undertaken to develop appropriate writing samples for assessment of composing tasks, to examine the utility and time requirements of scoring keys, particularly regarding their consistency and objectivity, and to study the efficacy of scoring guides. Five classes from second to sixth grade in a school in the Los Angeles Unified School District (LAUSD) participated in the study. The numerical equivalents assigned to the scores enabled staff to determine the appropriateness of the composing tasks on the basis of the students' success. Four of the five composing tasks, designed to be parallel to the LAUSD Surveys of Progress in Essential Skills, were appropriate. The scoring keys facilitated rapid scoring of the writing samples on a consistent and objective basis. Scoring the writing samples was not a time-consuming process, averaging less than 3 minutes per sample. These results confirm the feasibility of writing samples for large-scale assessment. (Author/CM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED220513

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

SWRL

SOUTHWEST REGIONAL LABORATORY
TECHNICAL NOTE

DATE: May 31, 1979

NO: TN 3-79-05

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

D. Grennell

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

WRITING SAMPLES, SCORING KEYS, AND SCORING GUIDES: A PILOT STUDY

Ann Humes

ABSTRACT

A pilot study of writing samples, scoring keys, and scoring guides at grades 2-6 is described, and the results are documented and discussed in terms of the purposes of the tryout. The study suggests that writing samples, scoring keys, and scoring guides are appropriate features of assessment instruments for writing.

TM 820 529

This document is intended for internal staff distribution and use. Permission to reprint or quote from this working document, wholly or in part, should be obtained from [illegible] 300 Lampson Avenue, Los Alamitos, California 90720

WRITING SAMPLES, SCORING KEYS, AND SCORING GUIDES: A PILOT STUDY

Ann Humes*

The validity of assessment instruments for writing that consist only of selected-response items is currently questioned by assessment authorities (e.g., Popham, 1979). Appropriate assessment of a student's writing ability should include a writing sample. Correspondingly, using appropriate writing-sample items and scoring keys for these items are primary considerations in constructing competency and proficiency assessment for writing skills.

Such writing sample items and scoring keys are included in the Los Angeles Unified School District's Surveys of Progress in Essential Skills (SPES). These assessment instruments are administered to students in grades one through six to determine students' competency on a number of basic skills in the areas of reading, math, and writing. Before item selection was finalized for the SPES, a pilot study was conducted of writing-sample items and their concomitant scoring keys and scoring guides. Items used in the pilot study parallel actual items in the SPES.

The purposes of the pilot study were (1) to determine the appropriateness of the composing tasks, (2) to examine the utility of the scoring keys, particularly in relation to their effecting consistency and thus objectivity in the scoring, (3) to determine the time requirements for scoring the writing samples when a scoring key is used, and (4) to study the efficacy of scoring guides as aids for evaluating writing samples. This report describes the pilot study and documents its results.

*Bruce Cronnell provided assistance with the data analysis.

PARTICIPANTS

The Los Angeles Unified School District (LAUSD) identified one of the schools within its boundaries for participation in the pilot study. The school is in a middle-income area. Five classes in the school were used in the pilot study, one at each of grades 2-6. The specific classes were selected by the school principal. Students were believed to be generally average to above average in scholastic ability. The classes included a relatively large bilingual population.

Two groups of writing-sample evaluators also participated in the pilot study. The first group (EG-1) consisted of four SWRL professional staff members of the Communications Skills Instructional Systems Capability Unit. The second group (EG-2) also consisted of four SWRL professional staff members: one from the Communications Skills Instructional Systems Capability Unit and three from other capability units.

MATERIALS

A writing-sample stimulus was prepared for each grade level. These stimuli were designed to parallel actual items written for the LAUSD Surveys of Progress in Essential Skills. The high degree of parallelism is evident in Figure 1, which gives an actual SPES item stimulus and its parallel item stimulus for the pilot study. In addition to the stimulus, each item included guidelines on which to write that were printed below the stimulus and on the reverse side of the paper.

Figure 1. Item from LAUSD/Survey of Progress in Essential Skills and Parallel Item for Pilot Study.

SPES ITEM:

Write a paragraph that tells about popcorn.

Tell how it looks.
Tell how it sounds when it is popping.
Tell how it smells.
Tell how it feels when you pick it up in your hand.
Tell how it tastes.

[Picture of Popcorn]

PILOT STUDY ITEM:

Write a paragraph that tells about an orange.

Tell how it looks.
Tell how it feels when you pick it up in your hand.
Tells how it smells when you open it to eat it.
Tell how it tastes.

[Picture of Orange]

Each writing stimulus was designed to elicit a specific discourse type appropriate for students in the corresponding grade level. Grade levels of the writing tasks were determined on the basis of data from the Proficiency Verification Systems for Language Arts (Humes, 1978).

The second-grade task elicited narrative discourse; students were to write sentences for pictures that together, in sequence, suggest a story. Descriptive discourse was evoked for the third-grade task, which asked students to describe a sensory object. The fourth-, fifth-, and sixth-grade items were designed to elicit

expository discourse: fourth grade, a newspaper account; fifth grade, map directions; sixth grade, a contrast of two elements of the same classification. The specific items for grades 2-6 are found in Appendix A.

The concomitant scoring key for each writing-sample item consisted of a performance/skill matrix. Performance ratings of good, average, and unacceptable are column heads, with checkmark blanks aligned opposite each skill criterion. Some criteria do not have blanks for all three possible ratings because one of the three is inappropriate for that criterion; e.g., "Shows originality" can be rated only "good" or "average" because an "unacceptable" rating for this criterion would unfairly assess beyond learned skills; "Indents first word of paragraph" has only good and unacceptable ratings because the student either does or does not indent.

The skills array of the key's matrix consists of content and form skills that are the component skills/subskills used to complete a specific composing task (e.g., "Orders information according to its importance" is a component skill of "Writes a newspaper story."). Correspondingly, these component skills produce the characteristics of good writing of a particular discourse type. One such characteristic of good writing is coherence. Coherence is not a measurable, discrete writing skill, but certain skills effect coherence in writing. For example, "Writes steps in chronological order" contributes to the coherence of a writing sample that explains how to do something. Thus this specific skill contributes coherence to a piece of expository writing.

A few of the criteria of some scoring keys were revised after the first group of evaluators scored the samples (see Results section). Both the original and revised forms are included in Appendix A with the items for the writing samples.

Scoring guides were developed to be used by the second group of evaluators. A scoring guide was written for each scoring key. The scoring guide provides guidelines on which to base a good/average/unacceptable score for each skill criterion on the scoring key. See Appendix B for an example of the scoring guides.

PROCEDURES

The pilot study was conducted within the first week of February, 1979. Materials were mailed to the school during the week before the tryout. Teachers were to administer the writing item at their convenience to their classes and then mail the writing samples back to SWRL. Teachers were also directed to allow students ample time to complete the writing samples, but in no case should that time exceed 45 minutes.

After all writing samples had been received, six samples from each grade level were selected for scoring. Samples were selected randomly with the exception of two excluding criteria: (1) Samples composed in a language other than English should not be selected; (2) although the scoring keys list legibility as an evaluated criterion, illegible samples should not be selected because their unreadability precludes evaluating student performance on the other criteria.

Each sample was coded by number. This number was to be recorded on the evaluator's scoring key along with his or her name and the date the sample was scored. The samples were circulated among evaluators by the author, who served as coordinator. The coordinator requested that evaluators record the time required to score each set of six samples. No guidelines were given to EG-1 for discriminating among good, average, and unacceptable performance ratings.

The samples and scoring keys were returned to the coordinator, who filed each set of scoring keys until she had scored that set of writing samples herself. Thus no evaluators saw ratings designated by another evaluator that might influence evaluator judgment.

When all evaluators had scored all samples for a grade level, numerical values were applied to the scoring categories: good = 4, average = 2, unacceptable = 0. Then scores were tallied, giving each writing sample four total scores, one from each evaluator. The data were then analyzed across samples and across evaluators.

On the bases of the analysis and evaluators' comments, some of the scoring keys were slightly revised. Then four of the five sets of writing samples were distributed for scoring by the second group of four evaluators. (Reasons for omitting one writing sample are given in the results section below.) All procedures but one were the same as for EG-1 above. The one varying procedure was that a scoring guide for each set of writing samples was given to each member of EG-2. The scoring guides suggested guidelines for determining the good/average/unacceptable ratings.

After the scoring process was completed, tables displaying the results of the data analysis were prepared.

RESULTS

As previously stated, the pilot study was designed to determine (1) the appropriateness of the composing tasks, (2) the utility of the scoring keys, particularly in relation to their effecting consistency and thus objectivity in the scoring, (3) time requirements for scoring writing samples when a scoring key is used, and (4) the efficacy of scoring guides as aids for evaluating writing samples.

Appropriateness of the Tasks

The numerical equivalents assigned to the scores (i.e., good = 4 points, average = 2 points, and unacceptable = 0 points) enabled staff to determine the appropriateness of the composing tasks on the basis of the students' success. Four of the five composing tasks were appropriate because nearly all samples received acceptable or better total scores. On those four sets, evaluators in EG-1 gave only five ratings, out of a possible 96, that totaled "unacceptable scores," for a 5% rate. Evaluators in EG-2, who received the scoring guides to facilitate evaluation, gave only one unacceptable total score on one sample for a 1% rate. The performance at each grade level is described below.

Grade 2. Although scores on the second grade writing samples were quite good, they may have been better if it hadn't been for ratings on the scoring-key criterion "Indents first sentence." This criterion

posed a scoring dilemma because of problems with the item format. The composition's title, which was printed on the page with the item, extended nearly all the way across the page. Students are accustomed to seeing text aligned to the left of the title, but the title placement allowed little space for both indenting the first line and still beginning to the left of the title. Consequently, indentation generally was rated as unacceptable. However, formatting of the parallel item on the actual SPES was revised to preclude this problem.

A total score of 32 was possible (four points for a good rating on each of eight criterion). Thus a score of 16 was the minimum score for an average rating for all criteria (two points for each average rating on each of eight criteria). Table 1 displays scores for the second-grade papers for each of the two groups of scorers. Accordingly, all students in the grade-two sample were able to perform this composing task successfully.

Grade 3. Since a score of 36 was possible for all criteria, a total score of 18 was the minimum score for an acceptable rating. One sample received unacceptable totals from two evaluators in EG-1. No sample received an unacceptable total from any member of EG-2. The second group's access to the evaluating guidelines in the scoring guide may account for these differences between EG-1 and EG-2. Table 2 shows the EG-1 and EG-2 scores for the third grade papers. These scores confirm the appropriateness of the task.

TABLE 1

a. EG-1 Ratings of Writing Samples for Grade 2 (Maximum possible score = 32)

Paper No.	Evaluator			
	A	B	C	D
1	28	28	26	26
2	32	26	20	24
3	28	24	20	26
4	28	28	20	24
5	32	28	32	26
6	32	24	28	28

b. EG-2 Ratings of Writing Samples for Grade 2 (Maximum possible score = 32)

Paper No.	Evaluator			
	A	B	C	D
1	26	28	28	26
2	16	24	24	24
3	24	28	28	26
4	22	22	26	22
5	28	26	30	32
6	20	18	30	26

TABLE 2

a. EG-1 Ratings of Writing Samples for Grade 3 (Maximum possible score = 36)

Paper No.	Evaluator			
	A	B	C	D
1	24	24	20	26
2	16	20	14	24
3	36	34	30	32
4	28	26	20	30
5	34	34	26	34
6	22	26	18	26

b. EG-2 Ratings of Writing Samples for Grade 3 (Maximum possible score = 36)

Paper No.	Evaluator			
	A	B	C	D
1	28	24	26	26
2	18	24	20	18
3	34	36	36	34
4	30	26	32	32
5	36	36	36	36
6	24	30	30	28

Grade 4. The fourth grade writing-sample item proved to be an inappropriate task. The item directed students to write a newspaper account from the listed "reporter's notes" (see Appendix A). Students received more unacceptable and marginal scores on this item than on any other composing task. The kinds of problems suggest that students may not be familiar with the important features of writing newspaper accounts despite the high frequency of this task in students' textbooks at both the third and fourth grades.

Rather than reorder the information according to its importance as the directions instructed them to do (a scoring criterion reflected this feature), students generally wrote the information in the same order as it was presented to them in the item. They also tended to repeat the language of the "reporter's notes" rather than use their own words. Consequently, evaluators gave students low ratings for "Rewrites rather than copies." Additionally, "Uses precise language" was not a relevant performance criterion because students were not using their own words.

Scores given by EG-1 would probably have been even lower if scoring standards had been given them: Two scorers perceived as "good" or "acceptable" an order of presentation that would not have been described as either good or acceptable in a scoring guide for the item.

Table 3 shows EG-1 scores for the fourth grade writing sample.

Because of the problems posed by this pilot-study item, the parallel item in the LAUSD Survey of Progress in Essential Skills was replaced by an item for writing a personal letter. Consequently, evaluators in EG-2 were not asked to score this set of writing samples.

TABLE 3

EG-1 Ratings of Writing Samples for Grade 4 (Maximum possible score = 40)

Paper No.	Evaluator			
	A	B	C	D
1	30	26	32	20
2	20	20	18	18
3	32	30	30	28
4	32	28	30	22
5	34	28	32	24
6	22	18	18	16

Grade 5. With a possible score of 40 for a good rating for all criteria, a total of 20 was the minimum acceptable score for an average rating on the fifth grade writing-sample item. Although revisions were made in the scoring criteria before EG-2 evaluated the papers (see subsequent discussion of the scoring keys), the number of criteria remained the same after the revision. One writing sample received an "unacceptable" total score from one evaluator in each group. Another writing sample received an "unacceptable" total score from two evaluators in EG-1 but none from EG-2. (The individual criteria eliciting the unacceptable ratings were not those revised on the scoring key before evaluation by EG-2.)

Table 4 displays the fifth-grade results for each of the two groups of evaluators.

TABLE 4

a. EG-1 Ratings of Writing Samples for Grade 5 (Maximum possible score = 40)

Paper No.	Evaluator			
	A	B	C	D
1	30	38	24	32
2	30	38	28	30
3	34	40	34	28
4	34	34	24	24
5	24	30	18	22
6	22	28	18	18

b. EG-2 Ratings of Writing Samples for Grade 5 (Maximum possible score = 40)

Paper No.	Evaluator			
	A	B	C	D
1	24	38	36	36
2	30	36	36	34
3	40	38	36	38
4	30	36	32	34
5	18	24	28	28
6	20	28	24	28

Grade 6. The total possible score differed for the two groups of evaluators because the scoring key was revised after EG-1 scored the samples. A score of 44 was possible for a good rating for all criteria using the original key given to EG-1, and a total score of 22 was the minimum score for an acceptable rating. However, a score of 56 was possible, with a concomitant total score of 28 as the minimum score for an acceptable rating for the revised key used by EG-2. Nevertheless, all students received acceptable or better total scores from all evaluators in both EG-1 and EG-2.

The scoring results for the sixth-grade assessment are shown in Table 5.

Utility of the Scoring Keys

In general, the scoring keys facilitated rapid scoring of the writing samples (see time section below) on a consistent and objective basis. While some evaluators were usually higher or lower than other evaluators, they tended to be fairly consistent within themselves--in giving consistently higher or lower scores. Thus while their scores may be higher or lower than those of other evaluators, the rank order of the samples generally matches that of other scores. Table 6 gives the rank ordering across evaluators on the samples for one grade level.

In most cases for which rank order varies, evaluator error is responsible. For example, one evaluator from EG-2 gave a disproportionately low score (in relation to other scorers) to a writing sample from the second grade. Examination of the specific writing sample and scoring key revealed that the scorer had given an unacceptable

TABLE 5

a. EG-1 Ratings of Writing Samples for Grade 6 (Maximum possible score = 44)

Paper No.	Evaluator			
	A	B	C	D
1	32	26	30	32
2	32	42	30	42
3	36	40	34	42
4	34	38	36	38
5	36	42	34	42
6	30	30	26	32

b. EG-2 Ratings of Writing Samples for Grade 6 (Maximun possible score = 56)

Paper No.	Evaluator			
	A	B	C	D
1	40	38	42	34
2	50	46	54	48
3	54	48	54	52
4	48	42	46	52
5	46	46	54	50
6	34	38	42	44

TABLE 6

Rank Order of Samples for Grade 3

a. EG-1 Rank Order of Samples

Paper No.	Evaluator			
	A	B	C	D
1	4	5	3.5	4.5
2	6	6	6	6
3	1	1.5	1	2
4	3	3.5	3.5	3
5	2	1.5	2	1
6	5	3.5	5	4.5

b. EG-2 Rank Order of Samples

Paper No.	Evaluator			
	A	B	C	D
1	4	5.5	5	5
2	6	5.5	6	6
3	2	1.5	1.5	2
4	3	4	3	3
5	1	1.5	1.5	1
6	5	3	4	4

rating for one criterion because the sentence for the picture was "missing." (It was not; the student had written a sentence about the picture.) Because the sentence was "missing," the scorer did not give any rating to the "orders sentences" criterion, although the sentences were in correct order. Thus the total score did not include the possible 6-8 points (out of 32) that were allowed by other evaluators.

In many cases of scorer error for EG-1, the error resulted from misinterpretation of the scoring key, which can be precluded by well written scoring guides. For example, ratings on the spelling criterion varied for EG-1; some evaluators scored students low for misspelling the same word many times, while other evaluators scored the repeated error as one misspelling. The scoring guide ameliorated this problem.

Corresponding with total score consistency, evaluators also tended to agree on the scoring of specific criteria on individual papers. (Agreement was determined by looking at the four scores given each paper by each evaluator group.) Nearly half of the time (47%) all four evaluators agreed completely; nearly half of the time (48%) they disagreed by only one degree (i.e., good/acceptable or acceptable/unacceptable); 36% of the time, three out of four agreed; 12% of the time, the scorers were equally split between scores. For only about 5% of the time did evaluators spread their scorers across the three degrees (i.e., good/unacceptable or good/acceptable/unacceptable). Again, the unacceptable scores in most such cases result from evaluator error.

Although the scoring keys did not generally pose problems, a few exceptions are described under the corresponding grade level below.

Grade 2. The only problem with the scoring key for grade two was the format problem explained above in the discussion of the tasks. Since the item rather than the scoring key required revision, the "Indents first sentence" criterion was not changed. Thus no revisions were necessary for grade two.

Grade 3. The scoring key for the third-grade sample posed no problems. Although sensory imagery criteria were reworded, the revision was not predicated on a scoring problem, but was intended to preclude one: The language of the revised version (see Appendix A) is more explicit.

Grade 4. Problems relating to the scoring key for the fourth-grade item were consequences of the item itself, as was previously discussed. No revisions were made in this scoring key because (1) the item was not scored by EG-2 and (2) the parallel item in the LAUSD SPES was replaced.

Grade 5. One content criterion was deleted from the scoring key for the fifth grade item before samples were scored by EG-2. "Apportions space devoted to a step according to its importance and complexity" was deleted because it was not applicable to the relatively short writing samples. Further, the relevance of this criterion for fifth grade may be questionable. One form-skills criterion, "Uses good sentence structure," was added for two reasons: (1) Evaluators in EG-1 felt the quality of the syntax should be considered, and (2) scorers

differed over which scoring criterion covered fragments and run-on sentences--punctuation or grammar. Adding the sentence-structure criterion ameliorated these problems.

Grade 6. The skill order was rearranged on the scoring key for grade six to facilitate rapid scoring of the samples; e.g., "Writes an appropriate title" was placed first to match reading order. In addition, three criteria were added. "Uses good sentence structure" was included for the same reasons as noted above for fifth grade. "Uses introductory sentence(s)" and "Shows originality" were added because scorers in EG-1 believed these criteria were needed to give students credit for including these features of good writing since they were not present in all samples. One criteria was reworded. "Arranges details in a logical order" was revised to a more precise statement: "Organizes the contrasts logically." Scorers in EG-2 had no difficulty with the revised scoring keys.

Time Required to Score Writing Samples

Scoring the writing samples was not a time-consuming process. Times recorded by evaluators confirmed this statement. Further, an accelerated rate of scoring can be anticipated for a larger set of writing samples because an evaluator's increasing familiarity with both the content elicited by the sample and the evaluation criteria of the scoring key can expedite scoring.

The mean time required to score writing samples is displayed in Table 7. Table 7 also indicates the range of time across evaluators.

TABLE 7

Time in Minutes Required to Score Samples

Grade Level	Mean	Range
2	1.95	1.17 to 2.67
3	2.25	1.5 to 2.5
5	2.93	2.5 to 3.33
6	2.8	1.67 to 3.33

The Efficacy of the Scoring Guides

As previously explained, scorers in EG-2 were given scoring guides to facilitate their decision making. To determine whether the benefits of scoring guides may be negated by the time required to read them, evaluators were requested to record preparation time. The average times for scorers to familiarize themselves with the scoring guide are the following:

Grade 2: Three minutes

Grade 3: Two minutes

Grade 5: Nine minutes

Grade 6: Five minutes

These averages suggest that the time required to study the scoring guide is not excessive if it facilitates more equitable and objective scoring. An indication of this equity and objectivity is suggested when a paper receives similar scores from multiple evaluators. Correspondingly, the range of scores across evaluators for each writing sample was generally lower for EG-2 than for EG-1 scores.

Table 8 gives the percentage range between highest and lowest scores for the evaluators of each group. In most cases the range between high and low scores is smaller for EG-2.

Table 8

% Range Between Highest and Lowest Scores

	% of Range Between Highest and Lowest Scores for Evaluators for Group 1	% of Range Between Highest and Lowest Scores for Evaluators for Group 2
Grade 2		
Paper 1	6%	6%
Paper 2	37%	25%
Paper 3	25%	13%
Paper 4	25%	13%
Paper 5	18%	18%
Paper 6	25%	37%
Grade 3		
Paper 1	17%	11%
Paper 2	28%	17%
Paper 3	17%	6%
Paper 4	28%	17%
Paper 5	22%	0%
Paper 6	22%	17%
Grade 5		
Paper 1	35%	35%
Paper 2	25%	15%
Paper 3	30%	10%
Paper 4	25%	15%
Paper 5	30%	25%
Paper 6	25%	20%
Grade 6		
Paper 1	14%	14%
Paper 2	27%	14%
Paper 3	18%	11%
Paper 4	9%	18%
Paper 5	18%	14%
Paper 6	14%	18%

CONCLUSIONS

The composing tasks developed for the LAUSD Survey of Progress in Essential Skills were appropriate with the exception of the item developed for fourth grade. The inadequacy of this task suggests that students are not so familiar with writing newspaper accounts as would be assumed from the presence of instruction on this skill in textbooks for grades three and four.

The scoring keys are useful in that they facilitate rapid scoring of the writing samples on a generally consistent and thus objective basis. Furthermore, they not only generate a total score to determine writing competency on the task, but also provide diagnostic information on the students' mastery of specific component skills or subskills relevant to a task.

Time required to score samples proved minimal enough to preclude "time constraints" as an argument against including composing task items in assessment instruments for writing. Furthermore, since the average time to score a sample from a set of six did not exceed three minutes in any instance, an accelerated scoring rate can be anticipated for a larger set because of familiarity with the content and the scoring criteria.

The use of scoring guides generally improves the consistency across evaluators of a sample, thus justifying the scorer's expenditure of the time necessary to familiarize himself or herself with the guide. However, if the classroom teacher is the only evaluator of the writing samples from his or her class, the time spent on studying a scoring guide may not be justifiable since individual scorers are usually consistent within themselves.

Finally, the pilot study results suggest something more important than the feasibility of writing samples for large-scale assessment: It suggests that, given an appropriate task and an evaluator using appropriate objective criteria, students really can write.

References

Humes, A. Proficiency verification systems (PVS): Skills indices for language arts (Technical Note No. 3-78-14). Los Alamitos, Calif.: SWRL Educational Research and Development, 1978.

Popham, W. J. Limited focus testing gets a test. Paper presented at AERA, San Francisco, April, 1979.

Appendix A

Writing-Sample Items and Scoring Keys

WRITING SAMPLE--GRADE 2

Read the title of the story. Look at the pictures that tell the story. Then write at least one sentence about the picture.

1.

2.

3.

The Bird Gets Away

SCORING KEY FOR GRADE TWO

Writing Sample - Grade 2
Paper No. _____
Evaluator _____
Date _____

DIRECTIONS FOR SCORING: Check the blank for the column that best describes the student's writing sample.

	Good	Average	Unacceptable
Includes sentence(s) for first picture.	—	—	—
Includes sentence(s) for second picture.	—	—	—
Includes sentence(s) for third picture.	—	—	—
Orders sentences correctly.	—	—	—
Writes legibly.	—	—	—
Capitalizes first word of sentence.	—	—	—
Includes periods at end of sentences.	—	—	—
Indents first sentence.	—		—

WRITING SAMPLE--GRADE 3

Write a paragraph that tells your class about an orange.

Tell how it looks.
Tell how it feels when you pick it up in your hand.
Tell how it smells when you open it to eat it.
Tell how it tastes.

--An Orange--

ORIGINAL SCORING KEY FOR GRADE 3

Writing Sample - Grade 3
Paper No. _____
Evaluator _____
Date _____

DIRECTIONS FOR SCORING: Check the blank for the column that best describes the student's writing sample.

	Good	Average	Unacceptable
Content			
Uses visual language (see).	—	—	—
Uses olfactory language (smell).	—	—	—
Uses tactile language (feel).	—	—	—
Uses gustatory language (taste).	—	—	—
Uses sentences that keep to the main idea.	—	—	—
Form			
Spells correctly.	—	—	—
Capitalizes first word of sentences and includes appropriate end punctuation.	—	—	—
Writes legibly.	—	—	—
Indents first word of paragraph.	—		—

REVISED SCORING KEY FOR GRADE 3

Writing Sample - Grade 3
Paper No. _____
Evaluator _____
Date _____

DIRECTIONS FOR SCORING: Check the blank for the column that best describes the student's writing sample.

	Good	Average	Unacceptable
Content			
Uses language that describes what an orange looks like.	—	—	—
Uses language that describes what an orange feels like.	—	—	—
Uses language that describes what an orange smells like.	—	—	—
Uses language that describes what an orange tastes like.	—	—	—
Uses sentences that keep to the main idea.	—	—	—
Form			
Spells correctly.	—	—	—
Capitalizes first word of sentences and includes appropriate end punctuation.	—	—	—
Writes legibly.	—	—	—
Indents first word of paragraph.	—		—

WRITING SAMPLE--GRADE 4

1. Read the notes that a reporter took on a science fair. The notes are *not* in order by what is most important.

2. Imagine that you are the reporter. Write a news article for your school newspaper about the science fair.

3. Remember:
   - Report all the important information.
   - Write the most important information first.

Reporter's Notes on the Science Fair

Fair last Friday, February 2

Fair held to decide on person to go to county science fair

Held on the playground

Ana Diaz, a fourth grader, won first place

All classes in school took part

Won for her experiment in the uses of electricity

Science Fair Results

SCORING KEY FOR GRADE 4

Writing Sample - Grade 4
Paper No. _____
Evaluator _____
Date _____

DIRECTIONS FOR SCORING: Check the blank for the column that best describes the student's writing sample.

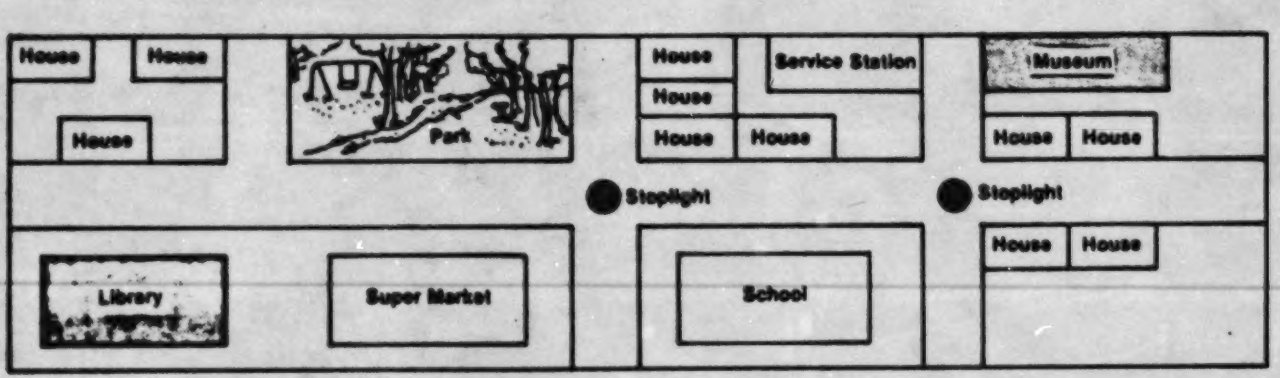
	Good	Average	Unacceptable
Content			
Includes all information in the notes.	—	—	—
Rewrites rather than copies.	—	—	—
Begins with the most important information.	—	—	—
Follows with less important information.	—	—	—
Uses precise language.	—	—	—
Uses appropriate transitional expressions.	—	—	—
Form			
Is grammatically correct.	—	—	—
Punctuates and capitalizes correctly.	—	—	—
Spells correctly.	—	—	—
Writes legibly, with appropriate margins, spacing, and indentation.	—	—	—

WRITING SAMPLE--GRADE 5

Study the map of the neighborhood. Find the library and the museum.

Write directions that tell how to get from the library to the museum. Write the directions in an order that is easy to follow.

Do not leave out any important information about which way to go or what to watch for along the way.



Directions to the Museum

ORIGINAL SCORING KEY FOR GRADE 5

Writing Sample - Grade 5
Paper No. _____
Evaluator _____
Date _____

DIRECTIONS FOR SCORING: Check the blank for the column that best describes the student's writing sample.

	Good	Average	Unacceptable
Content			
Includes all important information about landmarks.	—	—	—
Includes all important information about direction of movement.	—	—	—
Writes directions in a logical order.	—	—	—
Apportions space devoted to a step according to its importance and complexity.	—	—	—
Uses appropriate transitional expressions for sequence.	—	—	—
Uses precise and clear language.	—	—	—
Form			
Is grammatically correct.	—	—	—
Punctuates and capitalizes correctly.	—	—	—
Spells correctly.	—	—	—
Writes legibly, with appropriate margins, spacing, and indentation.	—	—	—

REVISED SCORING KEY FOR GRADE 5

Writing Sample - Grade 5
Paper No. _____
Evaluator _____
Date _____

DIRECTIONS FOR SCORING: Check the blank for the column that best describes the student's writing sample.

	Good	Average	Unacceptable
Content			
Includes all important information about landmarks.	—	—	—
Includes all important information about direction of movement.	—	—	—
Writes directions in a logical order.	—	—	—
Uses appropriate transitional expressions for sequence.	—	—	—
Uses precise and clear language.	—	—	—
Form			
Is grammatically correct.	—	—	—
Uses good sentence structure.	—	—	—
Punctuates and capitalizes correctly.	—	—	—
Spells correctly.	—	—	—
Writes legibly, with appropriate margins, spacing, and indentation.	—	—	—

WRITING SAMPLE--GRADE 6

Music and art are quite different school subjects.

- Write one or more paragraphs telling how they are different.
- You can tell about these differences:
  - You learn different things.
  - You study for them differently.
- Give your paper a title.

ORIGINAL SCORING KEY FOR GRADE 6

Writing Sample - Grade 6
Paper No. _____
Evaluator _____
Date _____

DIRECTIONS FOR SCORING: Check the blank for column that best describes the student's writing sample.

	Good	Average	Unacceptable
Content			
Contrasts facts about content learned.	—	—	—
Contrasts facts about study methods.	—	—	—
Applies all details to both music and art.	—	—	—
Arranges details in a logical order.	—	—	—
Uses appropriate transitional expressions for contrast.	—	—	—
Uses precise and clear langauge.	—	—	—
Writes an appropriate title.	—	—	—
Form			
Is grammatically correct.	—	—	—
Punctuates and capitalizes correctly.	—	—	—
Spells correctly.	—	—	—
Writes legibly, with appropriate margins, spacing, and indentation.	—	—	—

REVISED SCORING KEY FOR GRADE 6

Writing Sample - Grade 6
Paper No. _____
Evaluator _____
Date _____

DIRECTIONS FOR SCORING: Check the blank for the column that best describes the student's writing sample.

	Good	Average	Unacceptable
Content			
Writes an appropriate title.	—	—	—
Uses introductory sentence(s).	—	—	—
Contrasts facts about content learned.	—	—	—
Contrasts facts about study methods.	—	—	—
Organizes the contrasts logically.	—	—	—
Applies all details [facts] to both music and art.	—	—	—
Uses appropriate transitional expressions for contrast.	—	—	—
Uses precise and clear language.	—	—	—
Shows originality.	—	—	
Form			
Is grammatically correct.	—	—	—
Uses good sentence structure.	—	—	—
Punctuates and capitalizes correctly.	—	—	—
Spells correctly.	—	—	—
Writes legibly, with appropriate margins, spacing, and indentation.	—	—	—

APPENDIX B

Example Scoring Guide

SCORING GUIDE FOR GRADE TWO

Described below are guidelines for determining what constitutes a good, average, or unacceptable score on each criterion in the scoring key.

Includes sentences for first picture.

Good: One or two sentences tell about (1) the cat hiding in the bushes watching the bird it wants and (2) the bird drinking (or looking into) the water.

Average: One or two sentences are written, but they do not contain all the information noted above.

Unacceptable: The sentences do not pertain to the picture.

Includes sentences for second picture.

Good: One or two sentences tell about (1) the cat going after the bird and (2) the bird's starting to fly.

Average: One or two sentences are written, but they do not contain all the information noted above.

Unacceptable: The sentences do not pertain to the picture.

Includes sentences for third picture.

Good: One or two sentences tell about (1) the cat falling into the water and (2) the bird getting away from the cat.

Average: One or two sentences are written, but they do not contain all the information noted above.

Unacceptable: The sentences do not pertain to the picture.

Orders sentences correctly.

Good: All sentences are ordered in accordance with the sequence of the pictures.

Average: The sentences proceed sequentially from picture to picture, but one or two sentences are out of sequence.

Unacceptable: Many sentences are out of order so that no sequence is apparent in the story telling.

Capitalizes first word of sentences.

Good: All or most sentences are capitalized.

Average: Some sentences are capitalized.

Unacceptable: Many sentences are not capitalized.

Includes periods.

Good: Most or all sentences end with periods.

Average: Some sentences end with periods.

Unacceptable: Many or all sentences do not end with periods.

Writes legibly.

Good: All words are readable.

Average: Most words are readable and unreadable words do not interfere with comprehension of content.

Unacceptable: Many words are unreadable so that the content is difficult or impossible to read.

Indents first sentence.

Good: The first sentence of the story is indented.

Unacceptable: The first sentence of the story is not indented.